

VÖLCARE GMBH

KI-Beiträge transparent kennzeichnen

Leitfaden und Arbeitshilfe zur Differenzierung generativer KI-Nutzung in wissenschaftlichen, fachlichen und schöpferischen Arbeiten

Das Ampel-Modell ROT • GELB • GRÜN

ROT | R Substanz überwiegend durch generative KI hervorgebracht

GELB | G Menschliche fachliche Kernaussagen, KI unterstützend beteiligt

GRÜN | F Menschlicher Inhalt ausschließlich formal/sprachlich mit KI bearbeitet

Herausgeberin: VölCare GmbH

Verfasser: Chris Ola, im Auftrag der VölCare GmbH

Stand: 10. Oktober 2026 | Version 1.0 – redaktionell überprüfte Endfassung

Status: Fachliche Orientierungshilfe – freiwilliges Modell, keine Prüfungsordnung

Für Lernende, Lehrende, Prüfende, Bildungseinrichtungen, Unternehmen und weitere interessierte Stellen.

Publikations- und Verwendungshinweise

Chris Ola hat diesen Leitfaden im Auftrag der VölCare GmbH verfasst, die ihn herausgibt. Generative KI wurde bei der inhaltlichen Entwurfsarbeit und der sprachlichen Ausarbeitung in wesentlichem Umfang eingesetzt. Die VölCare GmbH hat die vorliegende Gesamtfassung einschließlich sämtlicher Anhänge inhaltlich und formal geprüft und zur Veröffentlichung freigegeben.

Selbsteinordnung der vorliegenden Arbeit nach dem Ampel-Modell: ROT. Wesentliche Teile der inhaltlichen Entwurfsarbeit und der Textausarbeitung wurden durch generative KI erbracht; die konzeptionelle Zielsetzung und die konkreten Anforderungen stammen aus menschlichen Vorgaben. Die Kennzeichnung macht die Entstehung transparent und ist kein Nachweis einer wissenschaftlichen Validierung des Modells. Eine Übernahme des Modells in verbindliche institutionelle Vorgaben erfordert eine eigenständige Prüfung der jeweils maßgeblichen rechtlichen und fachlichen Voraussetzungen.

Reichweite dieses Modells: Die Kennzeichnung beschreibt die Herkunft fachlich substanzieller Beiträge; sie ist keine Aussage über Wahrheit, Qualität, Rechtsverletzung, Täuschung oder persönliche Eignung. Vorrang haben jeweils geltende Rechtsvorschriften sowie verbindliche Prüfungs-, Publikations- und Organisationsvorgaben.

Zusammenfassung

Der vorliegende Leitfaden entwickelt ein freiwilliges Drei-Farben-Verfahren, um menschliche Leistungen und Beiträge generativer KI transparent und ohne vorschnelle Werturteile auseinanderzuhalten. ROT bezeichnet einen überwiegend KI-hervorgebrachten fachlich substanziellen Inhalt; GELB eine maßgeblich menschliche fachliche Leistung mit unterstützender KI-Formulierung, -Strukturierung oder begrenzter inhaltlicher Mitwirkung; GRÜN eine ausschließlich sprachliche oder formale KI-Überarbeitung eines bereits menschlich erstellten Inhalts. Ein vollständig ohne generative KI entstandener Abschnitt bleibt grundsätzlich unmarkiert; nicht verlässlich rekonstruierbare Vorgänge erhalten statt einer Scheinklassifikation einen neutralen Unsicherheitshinweis.

Maßgeblich ist nicht, wie viele Wörter die KI geschrieben hat, sondern wer die für das jeweilige Arbeitsergebnis entscheidenden Gedanken, Argumente, Bewertungen und Ergebnisse hervorgebracht hat. Die Kennzeichnung erfolgt nach möglichst kleinen inhaltlich selbstständigen Einheiten; gemischte Beiträge werden nach Möglichkeit aufgeteilt. Ein dokumentierter Entstehungsprozess unterstützt eine gewissenhafte Selbsteinstufung, beweist sie aber nicht zwingend.

Artikel 50 der Verordnung (EU) 2024/1689 in der durch Verordnung (EU) 2026/1744 geänderten Fassung regelt seit dem 2. August 2026 bestimmte Transparenzpflichten. Daraus folgt weder eine allgemeine Ampelkennzeichnungspflicht für Prüfungsarbeiten noch eine automatische Erlaubnis, in einem konkreten Prüfungsgang generative KI einzusetzen. Eine technische KI-Detektion kann Herkunft oder Täuschung nicht zuverlässig allein nachweisen. Der Leitfaden enthält deshalb ein regelgebundenes Prüfschema, Visualisierungsmuster, neun Fallbeispiele, eine Kurzanleitung und eine freiwillige Mustererklärung. [1–5, 8–12]

Inhalt

- 1 Ausgangspunkt und Gegenstand
- 2 Begriffe und fachliche Leitgedanken
- 3 Rechts- und Regelungsrahmen
- 4 Das Drei-Farben-Modell
- 5 Einstufungsverfahren und Entscheidungsschema
- 6 Unsicherheit, Mischfälle und Verzerrungen
- 7 Visuelle Umsetzung und Barrierefreiheit
- 8 KI-gestützte Selbsteinstufung und Datenschutz
- 9 Eigenleistung, Prüfung und die 50-Prozent-Frage
- 10 Empfehlungen für Anwendergruppen
- 11 Praxisfälle und Darstellungsmuster
- 12 Wissenschaftliche Einordnung und Evaluation
- 13 Schlussbetrachtung und Reichweite
- Anhang A Kurzanleitung und Prüfbogen
- Anhang B Freiwillige Mustererklärung
- Quellen- und Literaturverzeichnis
- Anhang C Einseitige Kurzfassung und Entscheidungshilfe

1 Ausgangspunkt und Gegenstand

Generative KI-Systeme können Ideen vorschlagen, Quellen erschließen, Textentwürfe erstellen, Argumente erweitern, bestehende Dokumente überarbeiten oder lediglich Rechtschreibung und Layout verbessern. Äußerlich ähnlich wirkende Ergebnisse können dadurch auf grundlegend verschiedenen Herstellungsprozessen beruhen. Für eine sachgerechte Einordnung sind mindestens die Herkunft der fachlichen Aussagen, die Herstellung der sprachlichen Fassung und die anschließende menschliche Kontrolle getrennt zu erfassen.

Transparenz kann für gute wissenschaftliche Praxis, faire Leistungsbewertung und nachvollziehbare redaktionelle Verantwortung hilfreich sein. Sie ist aber nicht mit einer pauschalen Pflicht gleichzusetzen, jedes technische Hilfsmittel offenzulegen. Der Anwendungsbereich der jeweiligen Regeln ist maßgeblich: Eine studentische Prüfungsleistung, ein Fachbeitrag, ein interner Arbeitsbericht und eine öffentlich verbreitete Information müssen nicht identischen Anforderungen unterliegen. Die DFG fordert in ihrem Förder- und Wissenschaftskontext eine transparente Offenlegung des eingesetzten generativen Modells sowie von Zweck und Umfang; das ICMJE formuliert für medizinische Fachpublikationen eigene Offenlegungserwartungen. Diese Beispiele belegen bereits existierende fachbezogene Anforderungen, nicht aber die allgemeine Verbindlichkeit eines einheitlichen Farbsystems. [8–10]

Der Leitfaden verfolgt drei Ziele: Er will erstens sinnvolle Differenzierungskategorien bereitstellen, zweitens deren Anwendung anhand überprüfbarer Entstehungshinweise ermöglichen und drittens eine lesbare, barrierearme Darstellung schaffen. Seine rot-gelb-grüne Systematik ist eine eigenständige konzeptionelle Empfehlung des Verfassers und der herausgebenden Organisation. Sie ist weder empirisch validierte Messskala noch unionsrechtlich vorgegebener Standard.

2 Begriffe und fachliche Leitgedanken

2.1 Vier Ebenen, die nicht verwechselt werden dürfen

Inhaltsherkunft beantwortet die Frage, wer die für einen Abschnitt maßgeblichen fachlichen Gedanken, analytischen Schritte, schöpferischen Entscheidungen und Schlussfolgerungen hervorgebracht hat. Für die Ampel

ist dies das vorrangige Kriterium. „Hervorbringen“ umfasst mehr als das bloße Einfügen oder Akzeptieren eines Vorschlags: Entscheidend ist, ob die wesentliche gedankliche Leistung bereits beim Menschen vorlag oder erst als neue, tragende Leistung aus einer KI-Ausgabe stammt.

Textproduktion bezeichnet die Ausformulierung, Satzbildung, sprachliche Verdichtung oder Übersetzung. Ein vollständig von einer KI formulierter Absatz kann im Modell GELB sein, sofern die fachlich maßgeblichen Inhalte im Ausgangsmaterial bereits menschlich bestimmt und in ihrer Bedeutung erhalten sind. Umgekehrt kann ein kurzer, von KI vorgeschlagener, aber entscheidender Lösungsansatz eine substanzielle maschinelle Leistung sein.

Redaktionelle Verantwortung bezeichnet die Prüfung, Auswahl, Freigabe und gegebenenfalls das Entstehen für veröffentlichte Aussagen. Sie bleibt von der Entstehungsherkunft zu unterscheiden: Die sorgfältige Prüfung eines KI-generierten Arguments beseitigt nicht rückwirkend dessen KI-Herkunft. Ebenso macht das bloße Abnicken von Vorschlägen nicht jede zuvor maschinell erbrachte fachliche Eigenleistung zur menschlichen.

Urheberschaft und prüfungsrechtliche Eigenleistung sind wiederum eigene Kategorien. Die tatsächliche Ideenherkunft beantwortet noch nicht die urheberrechtliche Frage nach einer schutzfähigen individuellen Gestaltung und auch nicht, welche persönliche Leistung eine konkrete Prüfung verlangt. [6, 7]

2.2 Arbeitsdefinitionen

„Generative KI“ bezeichnet hier Systeme, die auf Grundlage von Eingaben neu erscheinende sprachliche, visuelle oder andere Inhalte erzeugen oder verändern. Eine reine klassische Rechtschreibfunktion ohne generative Bearbeitung fällt nur dann unter dieses Modell, wenn tatsächlich generative KI an der Ausgabe beteiligt ist. Der Dienstname allein beweist keine bestimmte Bearbeitungstiefe.

„Fachlich substanziell“ sind Aussagen und Schritte, die das Verständnis, den Erkenntnisgewinn, die Interpretation, die Begründung, den Lösungsweg oder das Ergebnis einer Arbeit inhaltlich prägen. Dazu können die Auswahl entscheidender Tatsachen, die Entwicklung eines Untersuchungsdesigns, eine neue Argumentationslinie, eine fachliche Bewertung oder ein schöpferisch tragendes Konzept gehören.

„Überwiegend“ ist im Modell ein qualitatives, kein arithmetisches Merkmal. Es meint, dass KI-Beiträge die wesentlichen fachlich tragenden Elemente einer klar abgegrenzten Einheit prägen und die menschlichen Beiträge diese nicht in ihrer inhaltlichen Bedeutung übertreffen. Die Zahl der KI-formulierten Sätze ist hierfür nicht ausschlaggebend.

2.3 Gute Praxis statt Herkunftsvermutung

Die Quellenarbeit bleibt von der KI-Kennzeichnung getrennt. Ein wissenschaftlicher Sachverhalt benötigt fachlich belastbare Original- oder Sekundärquellen; ein KI-Dialog ersetzt diese nicht. Wird ein Textabschnitt mit ROT oder GELB markiert, ist damit weder die verwendete Quelle nachgewiesen noch der Text frei von fremden Rechten. Die in Publikationsrichtlinien geforderte menschliche Verantwortlichkeit besteht unabhängig von der Farbwahl fort. [9, 10]

3 Rechts- und Regelungsrahmen

3.1 EU-KI-Verordnung: geltende, aber begrenzte Transparenzpflichten

Die Verordnung (EU) 2024/1689 (KI-Verordnung bzw. AI Act), geändert durch die Verordnung (EU) 2026/1744 vom 8. Juli 2026 („Digital Omnibus on AI“), enthält in Artikel 50 eigenständige Transparenzpflichten für bestimmte Anbieter und Betreiber von KI-Systemen. „Anbieter“ und „Betreiber“ sind dabei unionsrechtliche Rollen, keine allgemeinen Synonyme für „Autor“ oder „Prüfling“. Für den konkreten Anwendungsbereich kommt es auf die Funktion und Verwendung des KI-Systems an. [1, 2]

Nach Artikel 50 Absatz 2 müssen Anbieter entsprechender Systeme grundsätzlich dafür sorgen, dass synthetisch erzeugte oder manipulierte Text-, Bild-, Audio- oder Videoausgaben maschinenlesbar markiert und als künstlich

erzeugt beziehungsweise manipuliert erkennbar sind, soweit dies technisch machbar ist. Die Vorschrift enthält insbesondere Ausnahmen für bloß assistierende Standardbearbeitungen oder Tätigkeiten ohne wesentliche Veränderung der Eingabedaten oder ihrer Semantik. Diese anbieterseitige technische Pflicht ist nicht identisch mit der hier vorgeschlagenen visuellen Einteilung ganzer Textpassagen. [1]

Artikel 50 Absatz 4 adressiert Betreiber bei bestimmten Deepfakes und bei mit KI erzeugten oder manipulierten Texten, die zum Zweck der Information der Öffentlichkeit über Angelegenheiten von öffentlichem Interesse veröffentlicht werden. Für letztere Textfälle besteht eine gesetzliche Ausnahme, wenn eine menschliche Überprüfung oder redaktionelle Kontrolle erfolgt ist und eine natürliche oder juristische Person die redaktionelle Verantwortung für die Veröffentlichung trägt. Diese Voraussetzungen sind konkret zu prüfen; die Ausnahme folgt nicht schon aus einer pauschalen Behauptung, man habe den Text gelesen. Artikel 50 Absatz 5 verlangt bei einschlägiger Transparenzinformation eine klare, unterscheidbare und zugängliche Mitteilung spätestens bei der ersten Exposition. [1]

Artikel 113 sieht für die hier behandelten Pflichten grundsätzlich die Anwendung seit dem 2. August 2026 vor. Anbieter bestimmter bereits vor dem 2. August 2026 in Verkehr gebrachter Systeme erhalten gemäß Artikel 111 Absatz 4 eine Übergangsfrist für Artikel 50 Absatz 2 bis zum 2. Dezember 2026; diese spezielle Übergangsregel darf nicht pauschal auf die Betreiberpflichten nach Artikel 50 Absatz 4 übertragen werden. Die Bestimmungen sind in der konsolidierten Fassung vom 27. Juli 2026 geprüft. [1, 2]

Die Europäische Kommission hat am 10. Juni 2026 einen freiwilligen europäischen Verhaltenskodex zur Transparenz KI-generierter Inhalte veröffentlicht und ergänzende Leitlinien beziehungsweise EU-Kennzeichnungssymbole bereitgestellt. Der Kodex soll die Erfüllung einschlägiger Artikel-50-Pflichten unterstützen; er ist nicht mit dem hier entwickelten Ampel-Modell gleichzusetzen. Wird Artikel 50 für eine konkrete Veröffentlichung relevant, muss dessen Rechtskonformität unabhängig von der Ampelfarbe geprüft werden. [3]

3.2 Urheberrecht: Werk, Person und Nutzung

Nach § 2 Absatz 2 UrhG sind Werke persönliche geistige Schöpfungen; § 7 UrhG bezeichnet den Schöpfer des Werkes als Urheber. Auch nach der Rechtsprechung des Gerichtshofs der Europäischen Union ist für den Werkbegriff die eigene geistige Schöpfung maßgeblich, die sich in freien und kreativen Entscheidungen ausdrückt. Für rein autonom maschinell erzeugte Inhalte ohne entsprechenden menschlichen Schöpfungsbeitrag fehlt nach diesen Grundsätzen regelmäßig eine unmittelbare menschliche Werkurheberschaft an der maschinellen Ausgabe. Bei komplexen Mensch-KI-Kooperationen hängt die Beurteilung dagegen vom konkreten menschlichen Gestaltungsspielraum und seiner erkennbaren Umsetzung ab. [6, 7]

Die Herkunft von Gedanken ist nicht mit dem urheberrechtlichen Schutz einer konkreten Ausdrucksform gleichzusetzen. Eine menschliche Idee kann unschutzzfähig sein, während eine konkrete schöpferische Ausgestaltung geschützt ist. Eine ROT-Markierung bedeutet deshalb nicht zwingend, das Endprodukt sei insgesamt urheberrechtlich ungeschützt. Ebenso garantiert GELB keinen Schutz oder die ausschließliche Urheberschaft des Erstellers.

Wer urheberrechtlich geschützte fremde Vorlagen bearbeitet, veröffentlicht oder verwertet, muss gegebenenfalls die Rechte Dritter beachten, insbesondere §§ 16, 19a, 23 UrhG. Eine transparente KI-Kennzeichnung ersetzt weder erforderliche Nutzungsrechte noch den wissenschaftlichen Quellenbeleg. Text- und Data-Mining-Regeln, etwa § 44b UrhG, sind nicht mit einer allgemeinen Erlaubnis zur beliebigen Übernahme geschützter Textpassagen gleichzusetzen. [6]

3.3 Schule, Hochschule, Prüfung und Weiterbildung

Welche KI-Unterstützung bei einer Leistungsbewertung zulässig ist, bestimmt sich nach der tatsächlich einschlägigen Rechts- und Regelungsebene: etwa Landesrecht, Ausbildungs- und Prüfungsordnungen, rechtmäßig konkretisierte Aufgabenstellungen, veröffentlichte Hilfsmittelvorgaben und gegebenenfalls wirksame vertragliche Bedingungen. Bei staatlichen Prüfungen treten grundrechtliche und verfahrensrechtliche Anforderungen hinzu; hochschulische Prüfungen richten sich insbesondere nach dem jeweiligen Hochschul- und Satzungsrecht.

Privatrechtlich organisierte Weiterbildungen dürfen nicht ohne gesonderte Prüfung wie staatliche Prüfungsverfahren behandelt werden.

Auch wenn eine Einrichtung eine einfache Offenlegungsregel verwendet, lässt sich daraus nicht ohne Weiteres eine allgemeine Rot-Gelb-Grün-Pflicht ableiten. Bestehen dagegen ausdrückliche Vorgaben, etwa eine andersfarbige Markierung sämtlicher KI-formulierter Passagen oder die Angabe von Werkzeug und Nutzungstag, haben diese im jeweiligen zulässigen Geltungsbereich Vorrang. Das Ampelsystem kann solche Vorgaben erläutern oder ergänzen, aber nicht eigenmächtig verändern.

Vor einer negativen prüfungsrechtlichen Konsequenz sind Tatbestand, Zuständigkeit, Verfahrensvorschriften und tatsächliche Feststellungen gesondert zu prüfen. Ob eine Entscheidung gebunden ist, einer fachlichen Beurteilung unterliegt oder einen rechtlich eröffneten Ermessensspielraum enthält, richtet sich nach der jeweiligen Norm. Ein farblicher Selbsteintrag ist weder ein normierter Täuschungstatbestand noch ein selbstständiger Bewertungsmaßstab.

3.4 Datenschutz, Vertraulichkeit und weitere Pflichten

Für personenbezogene Daten gelten insbesondere Artikel 5, 6 und – bei besonderen Kategorien wie Gesundheitsdaten – Artikel 9 DSGVO. Je nach Verarbeitung müssen Verantwortliche eine Rechtsgrundlage, Zweckbindung, Datenminimierung, technische und organisatorische Schutzmaßnahmen nach Artikel 32 sowie die gegebenenfalls erforderliche Auftragsverarbeitungsvereinbarung nach Artikel 28 DSGVO gewährleisten. Eine Datenübermittlung an externe KI-Dienste kann zusätzliche Anforderungen auslösen, etwa zu Drittlandtransfers gemäß Kapitel V DSGVO. Bloße Anonymisierungsbehauptungen reichen nicht aus, wenn Personen weiterhin identifizierbar sind. [11]

Unabhängig vom Datenschutz können Geheimhaltungs-, Arbeits-, Berufs-, Vertrags- und Geschäftsgeheimnispflichten einschlägig sein. Für sensible Unternehmens-, Gesundheits- oder Prüfungsunterlagen empfiehlt sich vorrangig eine datensparsame, institutionell freigegebene Lösung; ohne ausreichende Freigabe dürfen solche Inhalte nicht zur Einstufung an externe KI-Anbieter übertragen werden.

4 Das Drei-Farben-Modell

Die Ampel ist eine Herkunftskennzeichnung, keine Qualitätsampel. Rot steht nicht für „schlecht“, Grün nicht für „rechtlich erlaubt“ oder „moralisch gut“. Jede Einordnung bezieht sich auf eine benannte Aussageeinheit und den nach bestem Wissen nachvollziehbaren Erstellungsprozess. Das Modell gilt für Text und sinngemäß für fachliche Darstellungen, Tabellen, Grafiken und kreative Konzepte, soweit die Kategorien dort überhaupt sachgerecht angewendet werden können.

ROT | R Überwiegend KI-generierte fachlich substanzielle Beiträge

ROT ist anzuwenden, wenn die KI für eine eindeutig bestimmbare Einheit die fachlich tragenden Aussagen, wesentlichen Argumentationsschritte, Lösungen, Interpretationen oder schöpferischen Entscheidungen überwiegend hervorgebracht hat. Menschliches Prompting, Quellenprüfung, Auswahl einzelner Varianten und Schlussredaktion können umfangreich sein, ändern die ursprüngliche Herkunft der entscheidenden Substanz aber nicht automatisch. Eine spätere grundlegende menschliche Neuentwicklung kann demgegenüber zu einer neuen, anders zu prüfenden Fassung führen; bloße Bestätigung genügt nicht.

GELB | G Menschliche fachliche Kernaussagen mit unterstützender KI-Mitwirkung

GELB ist anzuwenden, wenn der Mensch die entscheidenden fachlichen Inhalte nachweisbar selbst entwickelt und die KI sie vor allem strukturiert, sprachlich ausformuliert, erläutert oder nur begrenzt ergänzt hat. Eine vollständige maschinelle sprachliche Fassung hindert GELB nicht, wenn ein ausreichend konkretes menschliches Konzept, fachliche Begründungen und Schlussfolgerungen bereits vorlagen und nachweisbar erhalten blieben. Allgemeine Themenangaben wie „Schreibe über Qualitätsmanagement“ genügen nicht: Erzeugt die KI daraufhin eigenständig die tragenden Argumente, ist ROT zu prüfen.

GRÜN | F Ausschließlich sprachlich-formale KI-Bearbeitung menschlicher Inhalte

GRÜN setzt einen bereits menschlich ausgearbeiteten inhaltlichen Bestand voraus. KI darf Grammatik, Rechtschreibung, Interpunktion, Lesbarkeit, Stil, Typografie oder reine Formatierung verbessern, ohne die fachliche Bedeutung, die Schlussfolgerung, die Stärke einer Aussage oder ihre argumentative Struktur materiell zu ändern. Eine KI-Korrektur von „möglicherweise“ zu „nachweislich“ oder die Ergänzung einer neuen rechtlichen Begründung ist nicht GRÜN, selbst wenn die Änderung unter dem Auftrag „nur sprachlich überarbeiten“ erfolgte.

Ohne KI oder ohne gesicherte Zuordnung. Vollständig ohne generative KI erstellte Einheiten bleiben im Normalfall unmarkiert; optional kann die Dokumentlegende ausdrücklich „unmarkiert = nach Erklärung keine generative KI“ vorsehen. Wenn Herkunft und Bearbeitung nicht belastbar bestimmbar sind, wird KEINE Ampelfarbe erzwungen. Stattdessen steht „Zuordnung offen“ mit kurzer Angabe des Grundes. Beide Fälle sind keine vierte Ampelfarbe.

4.1 Kompaktvergleich

Kategorie	Substanz	KI-Leistung	Typischer Fall
ROT / R	Tragende Inhalte überwiegend KI	Entwickelt maßgebliche fachliche Argumente	KI entwirft Analyse und Schlussfolgerung
GELB / G	Tragende Inhalte menschlich	Formuliert, gliedert, begrenzt ergänzt	Eigene fachliche Skizze wird KI-Text ergänzt
GRÜN / F	Menschlicher Inhalt unverändert	Korrigiert Stil, Orthografie, Layout	Vorhandener Bericht sprachlich geglättet
Unmarkiert	Vollständig menschlich	Keine generative KI in Einheit	Eigenständig erstellter Absatz
Hinweis „offen“	Unzureichend belegt	Entstehung nicht rekonstruierbar	Nicht dokumentierter Teamtext

5 Einstufungsverfahren und Entscheidungsschema**5.1 Prüfgegenstand und maßgebliche Belege**

Die kleinste sinnvoll abgrenzbare fachliche Einheit bildet den Ausgangspunkt: ein Absatz mit eigenständiger These, ein Teil einer Argumentation, ein Tabellenkommentar oder ein einzelnes Grafikelement. Eine pauschale Kennzeichnung des gesamten Dokuments ist nur geeignet, wenn diese Einstufung seine für die Aussage wesentlichen Teile tatsächlich zutreffend beschreibt; andernfalls ist eine abschnittsweise Darstellung vorzuziehen.

Zur Rekonstruktion kommen datierte Vorfassungen, Stichpunktkonzepte, Literatur- und Arbeitsnotizen, Promptverläufe, Änderungsvergleiche, Quellenvermerke und nachvollziehbare persönliche Arbeitsschritte in Betracht. Diese Materialien sind unterschiedlich aussagekräftig. Ein Prompt zeigt eine Anfrage, beweist aber nicht, dass sein Inhalt originär selbst erdacht war. Eine Versionshistorie belegt dokumentierte Änderungen, aber nicht

zwingend den Ursprung jeder darin enthaltenen Idee. Den tatsächlichen Erkenntnisweg muss der Ersteller nach bestem Wissen erklären.

5.2 Entscheidungslogik in sechs Schritten

Schritt 1 – Relevanz: Wurde in der betrachteten Einheit überhaupt generative KI verwendet? Falls nein, bleibt sie unmarkiert. Ist dies nicht aufklärbar, folgt ein neutraler Unsicherheitshinweis.

Schritt 2 – Ausgangsbestand: Lag der konkrete fachliche Gehalt vor der KI-Nutzung bereits in einer menschlichen Vorfassung vor? Welche Kernaussagen, Herleitungen oder Bewertungen sind darin nachweisbar?

Schritt 3 – Veränderungsprüfung: Hat die KI ausschließlich Stil oder Form geändert, ohne eine materielle inhaltliche Abweichung? Wenn ja, GRÜN. Bei jeder relevanten inhaltlichen Erweiterung ist GRÜN ausgeschlossen.

Schritt 4 – Substanzvergleich: Sind die entscheidenden fachlichen Aussagen und Begründungen weiterhin menschlichen Ursprungs und die KI-Leistungen lediglich stützend? Wenn ja, GELB. Entstanden dagegen die tragenden fachlichen Komponenten überwiegend durch KI, lautet die Einstufung ROT.

Schritt 5 – Mischfälle: Lassen sich materiell unterschiedliche Beiträge sinnvoll trennen, werden die betreffenden Passagen gesondert bezeichnet. Eine bloß formale KI-Bearbeitung nach früherer substanzieller KI-Generierung reduziert die rote Ursprungskennzeichnung nicht auf Grün.

Schritt 6 – Überprüfung: Gegenbelege, offene Varianten und wesentliche Unsicherheiten werden bewusst geprüft; die Einordnung wird gegebenenfalls korrigiert und die Begründung in knapper Form dokumentiert.

5.3 Entscheidungsschema auf einen Blick

Frage	Bei JA	Bei NEIN / unklar
Generative KI an der Einheit beteiligt?	Weiter zur Änderungsprüfung	Nein: unmarkiert / unklar: Hinweis „offen“
Bereits menschliche Substanz nur formal korrigiert?	GRÜN / F	Weiter zur fachlichen Herkunft
Fachlich tragende Substanz nachweislich menschlich?	GELB / G, falls KI unterstützt	Weiter zu KI-Substanzprüfung
Tragende Substanz überwiegend durch KI hervorgebracht?	ROT / R	Bei belegter Mischung aufteilen; sonst „offen“
Heterogene Beiträge abgrenzbar?	Kleinere Passagen getrennt markieren	Sorgfältig qualitativ abwägen; keine Scheinsicherheit

Prüfregel für neue Fassungen: Die Ampelfarbe ist keine lebenslange Eigenschaft einer Idee, sondern eine Erklärung über den konkreten Entstehungsweg des final dargestellten Inhalts. Ein nachträglich von einer Person substanziell neu entwickeltes Argument ist erneut anhand seines eigenen Entstehungsprozesses zu beurteilen. Reines Umarbeiten, Prüfen oder Umstellen von KI-Gedanken genügt regelmäßig nicht, um deren ursprünglichen fachlichen KI-Beitrag zu generieren.

6 Unsicherheit, Mischfälle und Verzerrungen

6.1 Vorrangregel für zusammengesetzte Passagen

Primär gilt die kleinräumige Zuordnung: Eine eigenständig entwickelte menschliche Problembeschreibung und eine darauf folgende KI-generierte Empfehlung erhalten getrennte Kennzeichnungen, wenn sie sich ohne Sinnverlust abgrenzen lassen. Bei einer heterogenen, nicht sinnvoll zerlegbaren Einheit geht die Herkunft ihrer fachlich entscheidenden Aussage gegenüber bloßen Wortanteilen vor. Ist diese entscheidende Aussage nachweislich

überwiegend KI-generiert, wird ROT verwendet und die gemischte Entstehung gegebenenfalls erläutert. Beruht die tragende Aussage auf menschlicher Leistung, ist GELB zutreffend, sofern KI inhaltlich nur dienend mitwirkte. Wenn beides nicht mit hinreichender Verlässlichkeit feststellbar ist, muss „Zuordnung offen“ statt einer erzwungenen roten oder gelben Farbe angegeben werden.

Die konservative Methode dient der Vermeidung einer untererklärten KI-Beteiligung, nicht der automatischen Belastung des Erstellers. Zweifel rechtfertigen keine frei erfundene Tatsachenfeststellung. Eine größere Abschnittseinheit darf auch mit dem erläuternden Hinweis „enthält getrennt dokumentierte KI-Beiträge“ versehen werden, solange die einzelnen Bestandteile nachvollziehbar markiert sind.

6.2 Ausschluss- und Gegenhypotheseprüfung

Die Selbsteinstufung beginnt mit einer vorläufigen Hypothese, beispielsweise GELB. Anschließend werden die stärksten Gegenargumente gezielt gesucht: Enthielt der menschliche Entwurf wirklich schon die entscheidende Begründung? Welche Textänderung geht über Sprachglättung hinaus? Könnte ein bestätigter KI-Vorschlag den Schluss erst ermöglicht haben? Lassen sich maßgebliche Unterschiede anhand der Versionen nachvollziehen? Erst danach wird die vorläufige Einordnung bestätigt, verändert oder als offen bezeichnet.

Solch ein Verfahren soll Bestätigungsfehler (confirmation bias), interessengetriebene Selbstdarstellung und unbewusste Bewertungsroutinen begrenzen. Aus der psychologischen Plausibilität möglicher Verzerrungen folgt jedoch kein Beweis, dass ein bestimmter Verfasser unzutreffend einstuft. Die Methode ist ein Transparenz- und Qualitätssicherungsinstrument, keine Glaubwürdigkeitsvermutung gegen Einzelpersonen.

6.3 Selbsteinstufung, Fremdeinschätzung und technische Hinweise

Eine Selbsterklärung besitzt eigenen Informationswert, bleibt aber eine nach bestem Wissen abgegebene Prozessdarstellung. Sie ist weder eine objektiv fehlersichere Inhaltsanalyse noch ein automatischer rechtlicher Beweis. Dieselbe Zurückhaltung gilt für Dritte, die lediglich den fertigen Text sehen: Ein gehobener Schreibstil, eine bestimmte Satzlänge oder formale Fehlerfreiheit sind kein zuverlässiger Herkunftsnachweis.

Untersuchungen von Weber-Wulff und anderen (2023) zeigten erhebliche Schwächen getesteter KI-Textdetektoren unter den damaligen Bedingungen. Liang und andere (2023) fanden in ihrer Stichprobe deutliche Fehlklassifikationen nicht muttersprachlicher englischer Texte. Die Studien belegen relevante methodische Risiken, aber keine für sämtliche späteren Detektoren unveränderlich geltende Fehlerquote. Insbesondere darf ein einzelner Detektorwert ohne belastbare weitere Tatsachengrundlage nicht als sichere Feststellung eines Täuschungsversuchs verwendet werden. [12, 13]

7 Visuelle Umsetzung und Barrierefreiheit

Für Word und PDF empfiehlt der Leitfaden eine zurückhaltende Farbhinterlegung oder eine farbige linke Randlinie, kombiniert mit einem ausgeschriebenen Kategorieetikett („ROT/R“, „GELB/G“, „GRÜN/F“) vor der betroffenen Passage. Werden lediglich farbige Unterstreichungen gewählt, können lange Fließtexte unruhig wirken; reine Randstriche sind hingegen dezent, müssen aber in Schwarz-Weiß-Versionen durch Textkennzeichen ergänzt werden. Farbige Rahmen eignen sich gut für einzelne längere Textblöcke. Eine Hintergrundfarbe sollte sehr hell sein, damit schwarzer Text kontrastreich bleibt.

Die Farbkodierung ist redundante visuelle Information, nicht der einzige Zugang. Diese Gestaltung folgt dem Zugänglichkeitsgrundsatz der WCAG 2.2, Erfolgskriterium 1.4.1 („Use of Color“); die Vorgaben zu Mindestkontrasten bei Text und visuellen Elementen sind je nach Medium zusätzlich zu prüfen. Eine automatische PDF-Konvertierung garantiert keine vollständig barrierefreie PDF/UA-Struktur. Formatvorlagen, echte Überschriften, lesbare Tabellen sowie manuelle Zugänglichkeitsprüfungen bleiben erforderlich. [14]

7.1 Legende und drucktaugliche Kennzeichnungsmuster

ROT | [R] Inhaltliche Substanz überwiegend durch KI hervorgebracht; Herkunft, nicht Qualität.

GELB | [G] Fachlich tragende Ideen menschlich; KI unterstützend für Text/Struktur/Inhalt.

GRÜN | [F] Ausschließlich formale oder sprachliche KI-Überarbeitung eines menschlichen Inhalts.

Im Schwarz-Weiß-Ausdruck bleiben die Buchstabenkürzel und ausgeschriebenen Kategorien sichtbar. Bei sehr kurzen Einheiten genügt beispielsweise „[G]“ unmittelbar vor dem Satz und die Legende am Dokumentbeginn. Bei umfangreichen Passagen empfiehlt sich ein umrahmter Abschnitt mit dem gut lesbaren Vorsatz „KI-Nutzung: GELB – unterstützende Ausformulierung“. Für digitale Publikationen sollten die Kategorieangaben im tatsächlichen Text und nicht ausschließlich als Kommentar, dekoratives Symbol oder Bild hinterlegt sein.

7.2 Muster einer markierten Fachpassage

GELB | G Beispiel einer unterstützenden KI-Ausformulierung

„Der Vorschlag zur Verbesserung des Einarbeitungsprozesses beruht auf der vom Verfasser dokumentierten Auswertung der eigenen Arbeitsabläufe; die KI hat die Darstellung sprachlich geordnet.“

ROT | R Beispiel eines substanziellen KI-Beitrags

„Die anschließende, eigenständig durch das KI-System vorgeschlagene Nutzwertanalyse wurde übernommen und nachträglich auf Plausibilität geprüft.“

GRÜN | F Beispiel ausschließlich formaler KI-Überarbeitung

„Der abschließende, bereits vollständig menschlich formulierte Reflexionsabschnitt wurde lediglich hinsichtlich Grammatik und Interpunktion korrigiert.“

Die Muster sind erfunden und veranschaulichen ausschließlich die Kennzeichnungslogik. Sie stellen keine tatsächlichen betrieblichen Vorgänge oder persönlichen Lernerfahrungen dar.

8 KI-gestützte Selbsteinstufung und Datenschutz

Eine generative KI kann die Selbsteinstufung sinnvoll vorbereiten, wenn ihr sowohl menschliche Ausgangsfassungen als auch KI-Vorschläge, Änderungen und die Endfassung rechtmäßig zugänglich gemacht werden. Sie kann Textvergleiche strukturieren, neu hinzugefügte Behauptungen aufzeigen, Sinnverschiebungen markieren und begründete Einstufungsoptionen liefern. Solche Analysen sind vor allem dann nützlich, wenn sie ausdrücklich auf vorhandene Versionen Bezug nehmen, statt vermeintliche Autorschaft allein aus dem Stil abzuleiten.

Das KI-System kann den inneren Entstehungsweg von Gedanken nicht unmittelbar beobachten und kann auch Vorfassungen missverstehen. Deshalb sind seine Aussagen als prüfbare Hypothesen zu behandeln. Die finale Einstufung erfolgt durch den verantwortlichen Menschen auf Grundlage des tatsächlichen Prozesses. Externe KI-Uploads erfordern zuvor eine Prüfung der Datenfreigabe, der Rechtsgrundlage, der Anbieterbedingungen, Aufbewahrung, möglicher Weiterverwendung und gegebenenfalls internationaler Datenübermittlungen. Bei vertraulichen oder personenbezogenen Materialien ist eine sichere lokale beziehungsweise freigegebene Verarbeitung vorzuziehen. [11]

Vorlage für eine zulässige KI-Vergleichsanfrage. „Vergleiche die beigefügten Versionen A (menschliche Vorfassung), B (KI-Ausgabe) und C (Endfassung). Benenne ausschließlich belegbare Änderungen der fachlich tragenden Aussagen; trenne formale Korrekturen, unterstützende Umformulierung und neu hinzugefügte Argumente. Gib für jede Passage Belege aus den Versionen und eine vorläufige Ampel-Empfehlung an. Wo der Entstehungsweg nicht feststellbar ist, schreibe ‚offen‘. Behaupte nicht, den Ursprung nicht dokumentierter Gedanken erkennen zu können.“

9 Eigenleistung, Prüfungen und die 50-Prozent-Frage

Bei einer bewerteten Leistung muss zwischen zugelassenen Hilfsmitteln, tatsächlich persönlich erbrachter fachlicher Leistung und der Bewertung des Arbeitsergebnisses unterschieden werden. Eine ROT-Kennzeichnung kann auf einen hohen maschinellen Beitrag zu prüfungsrelevanten Inhalten hinweisen. Ob diese Nutzung verboten, anzugeben, eingeschränkt zulässig oder vollständig erlaubt war, ergibt sich jedoch nicht aus der Farbe, sondern aus den konkret anwendbaren Regeln. Auch die Entdeckung einer nicht offengelegten Nutzung ersetzt die Feststellung der übrigen rechtlichen Voraussetzungen einer Sanktion nicht.

Ein Fall, in dem mindestens 50 Prozent der fachlich entscheidenden Bestandteile als ROT eingestuft werden, ist ein sinnvoller Anlass für eine vertiefte Prüfung der verlangten persönlichen Leistung, jedoch keine allgemeine rechtliche Schwelle. Weder der AI Act noch das Urheberrechtsgesetz formulieren für Prüfungsarbeiten eine solche Ampel-Prozentgrenze. Eine rein quantitative Zählung kann irreführen: Ein kurzer KI-generierter Schluss kann den gesamten Erkenntniswert prägen, während viele sprachlich KI-formulierte Wörter nur menschlich bereits festgelegte Inhalte transportieren. Eine Prozentangabe setzt außerdem eine offengelegte Bezugsgröße und Bewertungsmethode voraus.

Das Modell empfiehlt deshalb, Umfang und fachliche Tragweite getrennt zu beschreiben. Ist eine Zahl erforderlich, sollte dokumentiert werden, ob sie Seiten, Zeichen, Textblöcke oder zuvor definierte fachliche Inhaltseinheiten zählt; eine solche Zahl darf nicht nachträglich in einen normativen Zulässigkeitsmaßstab umgedeutet werden. Prüfende haben zunächst Regelungsgrundlage, Aufgabenstellung, erlaubte Hilfsmittel, tatsächlichen Entstehungsprozess und mögliche individuelle Verfahrensrechte zu ermitteln. ROT beweist für sich weder Täuschung noch Plagiat, Urheberrechtsverletzung oder das Nichtbestehen einer Prüfung.

10 Empfehlungen für Anwendergruppen

10.1 Lernende, Studierende und Verfasser

Prüfen Sie vor Beginn jeder bewerteten Arbeit die konkret geltenden KI-, Hilfsmittel- und Offenlegungsvorgaben. Bewahren Sie nach Möglichkeit menschliche Konzepte, Entwurfsstände und fachliche Notizen in angemessenem Umfang auf. Kennzeichnen Sie die tatsächliche Herkunft auch dann korrekt, wenn die KI lediglich Sätze ausformuliert hat; vermeiden Sie umgekehrt die nachträgliche Behauptung eigener fachlicher Gedankengänge, die tatsächlich erst durch KI entstanden sind.

Wenn eine eindeutige Einordnung misslingt, beschreiben Sie offen, welche Schritte gesichert und welche unklar sind. Eine sachlich begründete Unsicherheit ist redlicher als scheinpräzise Gewissheit. Vermeiden Sie das unzulässige Hochladen geschützter oder personenbezogener Materialien zur nachträglichen Beweissicherung.

10.2 Lehrende und Prüfende

Kommunizieren Sie die Bedeutung erlaubter KI-Unterstützung möglichst vor Bearbeitungsbeginn anhand konkreter Fallbeispiele. Trennen Sie die Herkunftskennzeichnung von der Bewertung persönlicher Kompetenz und von einer etwaigen Feststellung regelwidrigen Verhaltens. Prüfen Sie nicht allein den Prozentwert der Farbflächen, sondern die Rolle von KI für die fachlich entscheidenden Inhalte. Bei Zweifeln sollte eine transparente Sachverhaltsaufklärung vorgesehen werden, die die einschlägigen Verfahrensrechte beachtet.

Die von diesem Dokument empfohlenen Etiketten schaffen keine eigenständige Prüfungsbefugnis. In staatlichen und öffentlich-rechtlichen Verfahren ist insbesondere zwischen fachlicher Bewertung, gesetzlich gebundener Entscheidung und gegebenenfalls eröffnetem Verwaltungsermessen zu unterscheiden; der konkrete rechtliche Maßstab muss vorliegen.

10.3 Bildungseinrichtungen, Prüfungsträger und Stellen mit Aufsichtsaufgaben

Eine Organisation kann das Modell als Diskussionsgrundlage, Unterrichtsmaterial oder freiwillige Hilfestellung übernehmen. Beabsichtigt sie eine verbindliche Nutzung, muss sie deren Normgrundlage, Bestimmtheit, Gleichbehandlung, Zumutbarkeit, Bekanntgabe, Datenschutz und Übereinstimmung mit höherrangigem Recht eigenständig prüfen. Für die Bewertung ist festzulegen, was etwa „selbst erarbeitet“, „Textgenerierung“ und „reine Korrektur“ im konkreten Fach und Format bedeuten.

Denkbar sind anonymisierte Schulungsfälle, eine dokumentierte Bewertungsprobe mit mehreren Beurteilern und klare Rechtsbehelfs- beziehungsweise Klärungswege. Akkreditierungs- oder Aufsichtsstellen sollten nicht ohne eigene Zuständigkeit und normativen Bezug aus einem freiwilligen Ampelsystem verbindliche allgemeine Bewertungsmaßstäbe ableiten.

10.4 Unternehmen, Verlage und berufliche Teams

Für fachliche Berichte, Kundenkommunikation, Konzeptpapiere und öffentliche Beiträge bietet das Modell eine interne Herkunftstransparenz. Freigabezuständigkeiten, Qualitätskontrollen und Geheimhaltung sind getrennt festzulegen. Besonders bei Unternehmensveröffentlichungen über Angelegenheiten von öffentlichem Interesse ist unabhängig vom Modell zu prüfen, ob Artikel 50 Absatz 4 KI-Verordnung eingreift. Eine Herkunftskennzeichnung rechtfertigt weder die Verbreitung unrichtiger Tatsachen noch einen Verzicht auf fachliche oder datenschutzrechtliche Prüfung.

11 Praxisfälle und Darstellungsmuster

Die folgenden Sachverhalte sind ausschließlich illustrative, fiktive Beispiele. Ihr Zweck ist die konsistente Anwendung der vorstehenden Definitionen, nicht die Beschreibung realer Personen oder betrieblicher Vorgänge.

Fall	Sachverhalt	Einordnung und Grund
1	KI erhält nur ein Thema und entwickelt Gliederung, Thesen, Analyse und Empfehlung.	ROT: Die tragenden fachlichen Inhalte entstehen durch KI.
2	Ein Mensch notiert konkrete eigene Befunde, Argumentationsschritte und Schlussfolgerung; KI formuliert einen vollständigen Absatz.	GELB: Sprachfassung maschinell, fachliche Substanz menschlich.
3	Ein vollständiger menschlicher Bericht wird ausschließlich auf Rechtschreibung geprüft; Inhalt bleibt gleich.	GRÜN: Nur formale Korrektur.
4	Ein Absatz wird ohne generative KI recherchiert und verfasst.	Unmarkiert; optionale entsprechende Erklärung.
5	KI ersetzt in einer Korrektur „könnte“ durch „ist nachgewiesen“.	Nicht GRÜN: Aussageintensität materiell geändert; weitere Herkunftsprüfung nötig.
6	Menschlicher Befund wird von KI sprachlich dargestellt; eine neue tragende Ursachenhypothese stammt allein von KI.	Aufteilen: GELB für Befunddarstellung, ROT für Ursachenhypothese.
7	KI generiert fachliche Lösung; Mensch verifiziert Fakten und verbessert Stil.	ROT bleibt; redaktionelle Kontrolle ändert KI-Herkunft nicht.

Fall	Sachverhalt	Einordnung und Grund
8	Arbeitsgruppe übernimmt einen alten Text; Entstehungsprozess und KI-Anteil sind nicht rekonstruierbar.	„Zuordnung offen“, keine erzwungene Ampelfarbe.
9	KI gibt einzelne Stichworte; Mensch entwickelt daraus selbst eine neue eigenständige fachliche Argumentation.	Einzelfallprüfung: Bei nur unerheblichem KI-Impuls GELB möglich; bei tragender KI-Idee ROT für betroffenen Teil.

11.1 Mehrstufiger Sonderfall

Eine Verfasserin erstellt ein fachlich konkretes Exposé (menschliche Substanz). KI formt daraus eine ausführliche Darstellung (zunächst GELB). Anschließend fügt die KI eine bislang fehlende Risikobewertung mit eigenständiger Begründung hinzu (für diese Passage ROT). Die Verfasserin korrigiert die Rechtschreibung der gesamten Fassung (kein allgemeiner Farbwechsel zu GRÜN). Erfolgt zuletzt eine eigenständige fachliche Neubewertung durch die Verfasserin, wird deren konkreter Inhalt gesondert anhand seiner tatsächlichen Herkunft geprüft. Diese Abfolge zeigt, weshalb die Historie relevanter ist als der zuletzt verwendete Bearbeitungsbefehl.

12 Wissenschaftliche Einordnung und Evaluation

Das vorliegende Ampel-Modell ist eine normativ-praktische Klassifikation, keine psychometrisch validierte Skala. Wissenschaftlich gestützt sind insbesondere die Notwendigkeit menschlicher Verantwortlichkeit und transparenter Offenlegung in fachlichen Publikationskontexten (DFG, ICMJE, COPE), die Grenzen automatisierter Detektoren (Weber-Wulff u. a.; Liang u. a.) sowie zugänglichkeitsbezogene Gestaltungsgrundsätze (W3C). Aus diesen Quellen lässt sich jedoch nicht ableiten, dass exakt drei Farben und die hier vorgeschlagenen Grenzdefinitionen bereits empirisch als optimale oder allgemeingültige Lösung nachgewiesen wären. [8–10, 12–15]

Mögliche Stärken der Konzeption sind die Trennung von Gedankenherkunft und Formulierung, die niederschwellige optische Erkennung sowie die Unterscheidung zwischen dokumentierter Herkunft und Unsicherheit. Zu den Grenzen zählen unterschiedliche Auffassungen über die „tragende“ fachliche Leistung, nachträgliche Rekonstruktionsfehler, strategische Selbsteinstufungen, unklare Teambeiträge, Barrieren bei Farben und die von Fach zu Fach variierende Bedeutung schöpferischer Entscheidungen. Auch eine gute Dokumentation kann nicht alle mentalen Vorarbeiten objektiv sichtbar machen.

Für eine praktische Evaluation wären kontrollierte Vignetten mit bekannter Entstehungshistorie und unterschiedlichen Arbeitsgruppen sinnvoll. Prüfkriterien sollten unter anderem Verständlichkeit der Definitionen, Übereinstimmung unabhängiger Einstufungen, Fehlklassifikationen, benötigte Bearbeitungszeit, Sichtbarkeit in barrierearmen Formaten und wahrgenommene Fairness umfassen. Möglich wäre eine zweite Bewertungsrunde nach Schulung sowie die Prüfung, ob textliche Begründungen die Zuverlässigkeit erhöhen. Die hiermit vorgeschlagenen Erhebungen sind nicht durchgeführt worden; Ergebnisse oder Wirksamkeitswerte werden daher nicht behauptet.

Technische Herkunftsnachweise können das Modell ergänzen. Der C2PA-Standard dokumentiert bestimmte Herkunfts- und Bearbeitungsschritte von Medien mithilfe signierter Nachweisdaten; er beweist jedoch nicht umfassend, wessen Gedanken einen Text fachlich geprägt haben, und ist nicht mit einer menschlichen Eigenleistungsbewertung gleichzusetzen. Ebenso kann das Fehlen maschinenlesbarer Herkunftsdaten keinen sicheren Beweis ausschließlicher Menschlichkeit erbringen. [16]

13 Schlussbetrachtung und Reichweite

Eine tragfähige Offenlegung generativer KI-Nutzung muss das fachlich Wesentliche sichtbar machen, ohne rechtliche oder wissenschaftliche Gewissheit zu simulieren. Das Drei-Farben-Modell ordnet deshalb nicht Maschinenwörter, sondern den inhaltlichen Beitrag zu: ROT bei überwiegend KI-hervorgebrachter Substanz, GELB

bei menschlicher Kernaussage mit unterstützender KI-Nutzung und GRÜN bei ausschließlich formaler KI-Bearbeitung menschlicher Inhalte. Die nachvollziehbare Beschreibung gemischter oder offener Fälle ist Bestandteil, nicht Schwäche der Methode.

Der Leitfaden ist eine freiwillige fachliche Orientierungs- und Arbeitshilfe der VölCare GmbH. Er schafft durch seine Veröffentlichung keine allgemeinverbindliche Rechtsnorm, Prüfungsordnung, Beweisregel oder automatische Bewertung individueller Leistungen. Unabhängig hiervon können verbindliche rechtliche, satzungsrechtliche, vertragliche oder wirksam eingeführte institutionelle Vorgaben bestehen und eigene Rechtsfolgen auslösen. Die Zuständigen müssen jeden Sachverhalt nach den jeweils anwendbaren Regeln eigenständig prüfen. Die Ampel liefert dazu zusätzliche, gegebenenfalls hilfreiche Informationen, ersetzt aber weder sachliche Feststellungen noch fachliche, wissenschaftliche und rechtliche Verantwortung.

Anhang A Kurzanleitung und Prüfbogen

Der folgende Ablauf kann ohne weitere Unterlagen unmittelbar auf jeden abgrenzbaren Abschnitt angewendet werden. Er ist als individuelle Gedächtnisstütze konzipiert und begründet keine Pflicht, Arbeitsnotizen oder vertrauliche Entwürfe entgegen geltenden Vorgaben herauszugeben.

A.1 Geltende Vorgaben prüfen: Ist KI-Nutzung erlaubt, begrenzt, kennzeichnungspflichtig oder untersagt? Welche Veröffentlichungs- oder Prüfungsregeln gehen vor?

A.2 Einheit bestimmen: Welche einzelne These, Auswertung, Passage oder Grafik wird eingestuft?

A.3 Menschlichen Ausgangsbestand sichern: Welche fachlichen Aussagen standen vor dem KI-Einsatz konkret fest?

A.4 KI-Leistung vergleichen: Welche Aussagen wurden nur sprachlich überarbeitet, welche ergänzt oder eigenständig entwickelt?

A.5 Farbe auswählen: Nur Formänderung = GRÜN; tragende menschliche Substanz mit KI-Unterstützung = GELB; überwiegend KI-Substanz = ROT.

A.6 Mischfälle aufteilen; bei verbleibender Ungewissheit neutral „Zuordnung offen“ erklären.

A.7 Ausgabe zugänglich gestalten: Farbe plus [R], [G] oder [F] und Legende; keine reine Farbkommunikation.

A.8 Belege, Rechte, Datenschutz und fachliche Richtigkeit separat prüfen; erforderliche Tool-Offenlegungen ergänzen.

Kurzer dokumentierbarer Entscheidungsvermerk

Prüffrage	Freies Eintragungsfeld / Hilfstext
Abschnitt / Seitenbereich	Konkrete Passage benennen.
KI-Werkzeug und Nutzung	Name, ggf. Version und Datum nach tatsächlichem Kenntnisstand.
Menschliche Vorarbeit	Fachliche Kernaussagen und Beleglage beschreiben.
Maschinelle Änderung	Formulierung, Gliederung, neuer Inhalt oder anderes.
Ergebnis	ROT / GELB / GRÜN / unmarkiert / Zuordnung offen.
Einstufungsgrund	Ein bis drei nachvollziehbare vollständige Sätze.
Prüfung	Sach- und Quellenprüfung; ggf. offener Klärungsbedarf.

Anhang B Freiwillige Mustererklärung zur KI-Nutzung

Hinweis zur Verwendung: Die folgende Erklärung ist ein freiwilliges Muster. Sie muss an die tatsächliche Arbeit und die jeweils maßgeblichen Prüfungs-, Publikations- oder Beschäftigungsvorgaben angepasst werden. Das Muster ist weder eine eidesstattliche Versicherung noch ersetzt es eine verbindlich geforderte Eigenständigkeits- oder Hilfsmittelerklärung.

„Ich erkläre, dass ich die Beteiligung generativer KI an der Erstellung oder Bearbeitung der bezeichneten Arbeit nach bestem Wissen und anhand des mir nachvollziehbaren Entstehungsprozesses eingeordnet habe. Soweit das freiwillige Ampel-Modell verwendet wurde, kennzeichnet ROT fachlich überwiegend KI-hervorgebrachte Inhalte, GELB fachlich maßgeblich menschliche Inhalte mit unterstützender KI-Nutzung und GRÜN ausschließlich sprachlich oder formal KI-überarbeitete menschliche Inhalte. Nicht mit generativer KI bearbeitete Passagen sind unmarkiert, soweit die Legende nichts anderes bestimmt. Nicht zuverlässig rekonstruierbare Beiträge habe ich als solche kenntlich gemacht und nicht ohne hinreichende Grundlage einer Farbe zugeordnet.

Verwendete KI-Werkzeuge, ihre Einsatzgebiete und – soweit erforderlich oder bekannt – Nutzungszeitpunkte habe ich in einer ergänzenden Übersicht wahrheitsgemäß angegeben. Die fachliche Richtigkeit, die Beachtung der geltenden Quellen-, Datenschutz- und Nutzungsregeln sowie die Einhaltung der für meine Arbeit verbindlichen Anforderungen sind davon unabhängig zu prüfen. Die Kennzeichnung soll den tatsächlichen Arbeitsprozess transparenter machen; ich verbinde damit keine Behauptung, die Einordnung sei unfehlbar oder rechtlich verbindlich anerkannt.“

Ort, Datum: _____

Name / Unterschrift (falls erforderlich): _____

Empfohlene ergänzende Werkzeugübersicht

Werkzeug / Dienst	Zweck der Nutzung	Bereiche / Datum
[tatsächlichen Namen eintragen]	[z. B. sprachliche Überarbeitung; fachliche Recherche]	[Abschnitte / konkrete Angaben]
[bei Bedarf weitere Angaben]	[nur tatsächlich eingesetzte Funktionen]	[keine erfundenen Versionsnummern]

Quellen- und Literaturverzeichnis

Nachweissystem: Die eckigen Verweise [1] bis [16] in den Abschnitten beziehen sich auf die folgenden tatsächlich herangezogenen Rechts- und Fachquellen. Die Quellen tragen jeweils die im Text genannten Kernfeststellungen; das daraus entwickelte Ampel-Verfahren ist eine eigenständige Empfehlung. Abruf- und Bearbeitungsstand: 10. Oktober 2026.

- [1] Europäisches Parlament und Rat (2024/2026): Verordnung (EU) 2024/1689 vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz; konsolidierte Fassung vom 27. Juli 2026, insb. Art. 50, 111, 113.
<https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:02024R1689-20260727>
- [2] Europäisches Parlament und Rat (2026): Verordnung (EU) 2026/1744 vom 8. Juli 2026 (Digital Omnibus on AI), ABl. L 2026/1744 vom 24. Juli 2026. <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:32026R1744>
- [3] Europäische Kommission (2026): Code of Practice on Transparency of AI-generated Content; veröffentlicht am 10. Juni 2026; Website und Materialien des AI Office. <https://digital-strategy.ec.europa.eu/en/policies/code-practice-ai-generated-content>
- [4] Europäische Kommission / AI Act Service Desk (2026): Article 50 – Transparency obligations for providers and deployers of certain AI systems; Wiedergabe auf Grundlage der Fassung vom 27. Juli 2026.
<https://ai-act-service-desk.ec.europa.eu/de/ai-act/article-50>
- [5] Europäische Kommission (2026): Questions and answers – Code of Practice on Transparency of AI-Generated Content; insbesondere Zeitpunkte und Reichweite der freiwilligen Kodexpraxis. <https://digital-strategy.ec.europa.eu/en/faqs/code-practice-transparency-ai-generated-content>

- [6] Bundesrepublik Deutschland (2026): Gesetz über Urheberrecht und verwandte Schutzrechte (UrhG), insb. §§ 2, 7, 16, 19a, 23, 44b; konsolidierte Fassung auf „Gesetze im Internet“. <https://www.gesetze-im-internet.de/urhg/>
- [7] Gerichtshof der Europäischen Union (2019): Urteil vom 12. September 2019, C-683/17, Cofemel, ECLI:EU:C:2019:721, Rn. 29–32; ferner EuGH, Urteil vom 1. Dezember 2011, C-145/10, Painer, ECLI:EU:C:2011:798. <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:62017CJ0683>
- [8] Deutsche Forschungsgemeinschaft (2023): KI, ChatGPT und die Wissenschaften – DFG formuliert Leitlinien für Umgang mit generativen Modellen zur Text- und Bilderstellung, Pressemitteilung Nr. 39 vom 21. September 2023. <https://www.dfg.de/de/service/presse/pressemitteilungen/2023/pressemitteilung-nr-39>
- [9] International Committee of Medical Journal Editors (ICMJE, 2026): Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals, Fassung Januar 2026, Kap. „Use of Artificial Intelligence in Publishing“. <https://www.icmje.org/recommendations/>
- [10] Committee on Publication Ethics (COPE, 2024): Authorship and AI tools, COPE position, DOI 10.24318/cCVRZBms. <https://doi.org/10.24318/cCVRZBms>
- [11] Europäisches Parlament und Rat (2016): Verordnung (EU) 2016/679 (Datenschutz-Grundverordnung), insb. Art. 5, 6, 9, 28, 32 und Kapitel V. <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:32016R0679>
- [12] Weber-Wulff, D.; Anohina-Naumeca, A.; Bjelobaba, S.; et al. (2023): Testing of detection tools for AI-generated text. International Journal for Educational Integrity 19, Artikel 26. DOI 10.1007/s40979-023-00146-z. <https://doi.org/10.1007/s40979-023-00146-z>
- [13] Liang, W.; Yuksekgonul, M.; Mao, Y.; Wu, E.; Zou, J. (2023): GPT detectors are biased against non-native English writers. Patterns 4(7), 100779. DOI 10.1016/j.patter.2023.100779. <https://doi.org/10.1016/j.patter.2023.100779>
- [14] World Wide Web Consortium (W3C): Web Content Accessibility Guidelines 2.2, Understanding Success Criterion 1.4.1 „Use of Color“ und 1.4.3 „Contrast (Minimum)“. <https://www.w3.org/WAI/WCAG22/Understanding/use-of-color.html>
- [15] Miao, F.; Holmes, W. (2023): Guidance for generative AI in education and research. UNESCO, Paris. <https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>
- [16] Coalition for Content Provenance and Authenticity (C2PA): C2PA Specifications, Version 2.2, „Content Credentials“ und Informationen zur Medienprovenienz. <https://spec.c2pa.org/specifications/specifications/2.2/index.html>

KI-Mitwirkung auf einen Blick

Das freiwillige Ampel-Modell ROT • GELB • GRÜN

DAS ENTSCHEIDENDE KRITERIUM

Wer hat die fachlich tragenden Gedanken, Begründungen und Ergebnisse hervorgebracht?

Nicht ausschlaggebend sind die Zahl KI-formulierter Wörter, das bloße Prompting oder die spätere menschliche Endkontrolle.

01 DIE DREI KATEGORIEN *Herkunftskennzeichnung • keine Qualitätsampel*

ROT [R]	KI prägt die fachliche Substanz Die KI hat tragende Argumente, Bewertungen, Lösungen oder Schlussfolgerungen überwiegend hervorgebracht.	BEISPIEL KI entwickelt aus einer bloßen Themenvorgabe Analyse und Empfehlung.
GELB [G]	Fachliche Kernaussagen stammen vom Menschen Ein konkretes menschliches Konzept trägt den Inhalt; KI formuliert, gliedert oder ergänzt nur begrenzt.	BEISPIEL KI verfasst den Absatz aus eigenen dokumentierten Befunden und Schlussfolgerungen.
GRÜN [F]	Nur Sprache oder Form mit KI bearbeitet Bereits menschlich ausgearbeiteter Inhalt bleibt fachlich unverändert; KI korrigiert nur Stil, Sprache oder Layout.	BEISPIEL KI korrigiert Rechtschreibung eines fertigen, menschlich erstellten Berichts.

MERKE Auch ein vollständig KI-formulierter Text kann GELB sein, wenn die tragende fachliche Leistung bereits menschlich vorlag. Rein formale Nacharbeit macht zuvor KI-generierte fachliche Substanz nicht GRÜN.

02 ENTSCHEIDUNGSPFAD *Für jede fachlich selbstständige Einheit gesondert anwenden*

VORAB Eine eigenständige Aussageeinheit abgrenzen; trennbare Mischfälle getrennt markieren.			
1	Generative KI bei der Entstehung beteiligt?	JA Weiter zu 2 ↓	NEIN unmarkiert • unklar: „Zuordnung offen“
2	Bereits menschlicher Inhalt nur sprachlich/formal bearbeitet, ohne fachliche Änderung?	JA GRÜN [F]	NEIN Weiter zu 3 ↓
3	Entscheidende fachliche Aussagen menschlich; KI nur stützend oder begrenzt ergänzend?	JA GELB [G]	NEIN Weiter zu 4 ↓
4	Tragende fachliche Substanz überwiegend von KI hervorgebracht?	JA ROT [R]	NICHT KLÄRBAR „Zuordnung offen“

MISCHFÄLLE & ZWEIFEL Nicht trennbare Beiträge nach Herkunft der fachlich entscheidenden Aussage einordnen. Ist diese nicht verlässlich feststellbar, den Grund der offenen Zuordnung nennen; keine Farbe erzwingen. Vorfassungen und Gegenbelege prüfen.

03 GRENZEN & VERANTWORTUNG

NICHT AUS DER FARBE ABLEITBAR Die Ampel bewertet weder Richtigkeit oder Qualität noch automatisch zulässige KI-Nutzung, persönliche Prüfungsleistung, Täuschung oder Urheberrechtsverletzung.	WEITERHIN EIGENSTÄNDIG PRÜFEN Verbindliche Rechts-, Prüfungs- und Publikationsvorgaben gehen vor. Quellen, fachliche Richtigkeit, Rechte, Datenschutz und Freigabe bleiben zu prüfen; ein KI-Detektor allein beweist keine Herkunft.
---	--

FREIWILLIGES MODELL • Keine allgemeinverbindliche Ampelpflicht und keine Prüfungsordnung.

Herausgeberin: VölCare GmbH • Grundlage: Leitfaden von Chris Ola im Auftrag der VölCare GmbH • Stand: 10.10.2026 / Version 1.0.

Transparenz: Der Leitfaden stuft seine KI-Mitwirkung als ROT ein; diese Kurzfassung wurde mit generativer KI verdichtet und gestaltet. Details und Quellen: Langfassung.